

## Ders Yönetim Sistemi Yazılımı Kullanıcı Etkileşimlerinin Sınıflandırma Algoritmaları ile Analizi (\*)

Hakan GÜLDAL (\*\*)

Yılmaz ÇAKICI (\*\*\*)

**Öz:** Bu araştırmanın amacı, Ders Yönetim Sistemi Yazılımı (DYS) kullanan öğrencilerin sisteme olan etkileşimlerine dayalı olarak veri madenciliği sınıflandırma algoritmaları ile öğrencilerin akademik başarılarının ne ölçüde tahmin edilebileceğini ve onların akademik başarılarının tahmin edilmesinde algoritmalara hangi etkileşimlerin bilgi kazancı sağladığını belirlemektir. Araştırma verileri, Trakya Üniversitesi Eğitim Fakültesinde 2015-2016 eğitim öğretim yılında öğrenim görmekte olan 70 öğrencinin Bulutders DYS yazılımını kullanma sürecindeki etkileşimlerinden elde edilmiştir. Veri madenciliği sınıflandırma algoritmaları olarak Naive Bayes, Karar Ağacı (C4.5) ve En Yakın k-Komşu algoritmaları kullanılmıştır. Araştırma sonucunda elde edilen bulgulara göre, sınıflandırma algoritmalarının doğruluk değerlerinin %55.7 ile %64.3 arasında olduğu ve DYS yazılımındaki ödevlerden elde edilen öğrenci etkileşimlerinin en yüksek bilgi kazancını sağladığı ortaya çıkmıştır.

**Anahtar Kelimeler:** Eğitsel Veri Madenciliği, Öğrenme Analitikleri, Sınıflandırma Algoritmaları

### Analysis of Course Management System Software Users' Interactions Using Classification Algorithms

**Abstract:** The aim of this study is to determine the level of information gain obtained through the students' interactions of the Course Management System (CMS) software, and how well students' academic achievement can be predicted with the use of data mining classification algorithms, and also which interactions provide information gain to algorithms in predicting academic achievement. Research data were obtained through the interactions of seventy students who are using Bulutders CMS Software, and who are studying in the Faculty of Education, Trakya University in 2015-2016 academic term. Naive Bayes, Decision Tree (C4.5) and k-Nearest Neighbour algorithms were used as data mining classification algorithms. The research findings revealed that accuracy values of classification algorithms were 55.7 and 64.3 percent, and the student interactions obtained from assignments in CMS software provided the highest information gains.

**Keywords:** Educational Data Mining, Learning Analytics, Classification Algorithms

**Makale Geliş Tarihi: 08.10.2017**

**Makale Kabul Tarihi: 29.11.2017**

---

\*) Bu çalışmanın kısa bir özeti "International Computer and Instructional Symposium-ICITS2017 ,Mayıs 2017,Malatya/Türkiye" sempozyumunda sunulmuştur.

\*\*) Öğr.Gör.Dr,Trakya Üniversitesi Eğitim Fakültesi BÖTE Bölümü (e-posta: hguldal@trakya.edu.tr)

\*\*\*) Doç.Dr. Trakya Üniversitesi Eğitim Fakültesi Fen Bilgisi Eğitimi Anabilim Dalı (e-posta: yilmazcakici@trakya.edu.tr)

## I. Giriş

Günümüzde kullanıcılar ve cihazlar tarafından farklı formatlarda ve çok büyük miktarlarda veri üretilmekte ve bu üretilen veriye çok hızlı bir şekilde ulaşılabilmektedir. Sosyal paylaşım platformlarında paylaşılan fotoğraflar, hareketli görüntüler, duygu durumları, düşünceler, beğeniler, blog sistemlerinde yer alan içerikler, arama motorlarında kullanılan anahtar kelimeler, alışveriş sitelerinden satın alınan veya yorum yapılan ürünler, web sitesi gezintilerindeki kayıt bilgileri, tıklama eylemleri vb. bilgisayar ve yazılım sistemleri kullanılarak küresel ölçekte, kesintisiz bir şekilde ve çok büyük miktarlarda veri üretilmektedir. Üretilen bu büyük veri yığınının içerisinde kullanışlı ve anlamlı bilgilerin ortaya çıkarılması özellikle son dönemlerde önem kazanmaya başlamıştır. Bilgisayar bilimleri içerisinde yer alan ve bu amaca odaklanmış bir çalışma alanı olan veri madenciliği, makine öğrenmesi, örüntü tanıma, istatistik, olasılık teorisi, veri görselleştirmesi gibi alanlardaki teknikleri bir arada kullanan disiplinler arası bir alandır (Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1998). Veri madenciliği, veritabanı gibi yapısal veriler veya web sayfası, düz metin gibi yapısal olmayan üzerinde sınıflandırma, kümeleme, birliktelik kuralları gibi işlemler için kullanılır.

Veri madenciliği ile birlikte verilerden anlamlı bilgiler ve tahminler çıkarmayı amaç edinen diğer bir disiplin de istatistiktir. Bu ortak amaca rağmen, veri madenciliği analizleri ile istatistik analizleri arasında bazı yönlerden farklılıklar bulunmaktadır. İstatistiksel analizler belirli bir hipotezle başlar ve analiz sonucunda hipotez kabul edilir ya da reddedilir. Veri madenciliği ise bir hipoteze gerek duymaz dolayısıyla analiz sonucunda kabul veya red durumu söz konusu değildir. Eldeki veriyi analiz eder ve içerisindeki anlamlı örüntüleri ortaya koyar. İstatistiksel analizler tüme varımsal niteliktedir, veri madenciliği analizleri ise tümden gelimseldir. İstatistik analizler genellikle ana kütlelerin bir kesiti alınarak gerçekleştirilir. Veri madenciliği araştırmalarında verilerin tümü analizlerde kullanılır (Tüzüntürk, 2010).

Veri madenciliği, sağlık (Ginsberg et al., 2009) biyoloji (Baldi & Brunak, 2001) astronomi (Fayyad, Weir, & Djorgovski, 1993) gibi pek çok alanda kullanılmaktadır. Eğitim alanında ise, eğitsel veri madenciliği ve öğrenme analitikleri gibi araştırma alanlarında, sınıflandırma, kümeleme, birliktelik kuralları gibi yöntem ve teknikleri, bireysel özelliklerine göre benzer öğrencilerin gruplanması, öğrenci ilgi alanlarının ve eğilimlerinin tahmini, öğrenme içeriklerinin sunulması, düşük motivasyona sahip öğrencilerin tespiti, kavram yanlışlarının ortaya çıkarılması gibi çeşitli görevler için kullanılabilir.

Eğitim alanında gerçekleştirilen veri madenciliği çalışmalarında gerekli olan veri setini oluşturmak için öğrenme yönetim sistemleri, sosyal paylaşım siteleri, wikiler, tartışma panoları gibi yazılımlardan elde edilen veriler kullanılabilir. Bu amaçla kullanılacak yazılımlardan biri de, Ders Yönetim Sistemi (DYS) yazılımlarıdır. DYS yazılımları, geleneksel ortamlarda yüz yüze gerçekleştirilen eğitim öğretim faaliyetlerini desteklemek amacıyla kullanılırlar. DYS yazılımlarıyla öğretmenler verdikleri derslerle ilgili duyuru ve dokümanları çevrimiçi olarak yayına koyarak öğrencileriyle paylaşabilir, ödevler verebilir ve alabilir, öğrencilerin çalışmalarını çevrimiçi olarak takip edebilir ve

değerlendirebilirler aynı zamanda gerçek zamanlı mesajlaşma, tartışma panoları, forum gibi araçları kullanarak birbirleriyle iletişim kurabilirler (Martins & Kellermanns, 2004). Veri madenciliği yönünden bakıldığında, DYS yazılımı üzerinde gerçekleştirilen; sisteme giriş, tıklama gibi her bir öğrenci etkileşiminin kayıtlarının tutulması, araştırmalarda gerekli olan veri setinin kaynağını oluşturmaktadır.

Eğitim alanında veri madenciliği yöntem ve teknikleri, öğrencilerin akademik performanslarının modellenmesinde de kullanılmaktadır. Akademik performans modellenmesinde; öğrenme yönetim sistemi vb. ortamlarda üretilen verilerle öğrencilerin akademik performansları veri madenciliği yöntem ve teknikleri ile modellenmeye çalışılır. Literatürde, akademik performans modellenmesiyle ilgili gerçekleştirilmiş çok sayıda araştırma bulunmaktadır (Buniyamin, Mat, & Arshad, 2016), (Bhardwaj, 2011), (Christian & Ayub, 2014)(Mishra, Kumar, & Gupta, 2014), (Shahiri, Husain, & Rashid, 2015), (Márquez-Vera, Cano, Romero, & Ventura, 2013), (Del, 2016), (Ahmad, Ismail, & Aziz, 2015), (Baradwaj & Pal, 2012), (Bhardwaj, 2011) (Akcapinar, Altun, & Aşkar, 2015), (Osmanbegović & Suljić, 2012), (Akçapınar, Coşgun, & Altun, 2013) (Márquez-Vera et al., 2013).

#### **A. Araştırmanın Amacı**

Bu araştırmada, aşağıdaki sorulara cevap bulunması amaçlanmıştır;

(1) DYS yazılımı kullanan öğrencilerin sistemle olan etkileşimlerinden elde edilen verilerden, veri madenciliği sınıflandırma algoritmalarıyla akademik başarıları ne ölçüde tahmin edilebilir?

(2) Öğrencilerin akademik başarılarının tahmin edilmesinde, sınıflandırma algoritmalarına DYS yazılımı ile olan hangi etkileşimleri ne kadar bilgi kazancı sağlamaktadır?

#### **II. Materyal Metot**

##### **A.DYS Yazılımı**

Araştırmada DYS olarak araştırmacılar tarafından geliştirilen Bulutders yazılımı (Güldal, 2015) kullanılmıştır. Bulutders, bulut platformu üzerinde web tabanlı olarak çalışan, kullanım ve işleyiş şekliyle kullanıcıların aşına oldukları sosyal ağ uygulamalarına oldukça benzeyen bir DYS yazılımıdır. Bulutders yazılımında, sınırsız sayıda ders ve sınırsız sayıda kullanıcı tanımlanabilmekte, kullanıcılara ait bilgiler, ders materyalleri, duyurular vb. bulut platformu üzerinde tutularak, bu verilere herhangi bir yerden, herhangi bir zamanda ve herhangi bir cihazla erişim sağlanabilmektedir. Bulutders üzerinde tanımlanan dersler “Herkes Açık” ve “Erişim Kodu İle” olmak üzere iki erişim şekline sahiptir. Herkes açık olarak tanımlanan bir derste öğrenciler istedikleri herhangi bir dersi herhangi bir onay koşulu olmaksızın takip edebilmekte ve sistem üzerinde bu ders içinde yer alan tüm veriler öğrenci için erişilebilir hale gelmektedir. Kod kullanılarak erişimde ise öğrenciler sistem tarafından üretilen ve ders sorumlusu tarafından sağlanan anahtar bir erişim kodunu kullanarak sistem üzerinde bir derse katılabilmektedir. Bu araştırmada, İnternet Tabanlı Programlama Dersine öğrencilerin yetkisiz erişimleri engellemek için ders sorumlusu tarafından sağlanan erişim kodu tercih edilmiştir.

### B.Kullanıcılar

Araştırmada kullanılan veriler 2015-2016 eğitim öğretim yılında Trakya Üniversitesi Eğitim Fakültesi Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümünde öğrenim görmekte olan ve İnternet Tabanlı Programlama dersini alan 70 öğrenciden elde edilmiştir. İnternet Tabanlı Programlama dersi, Bulutders DYS yazılımı üzerinde tanımlanmış, derse ait duyurular, materyaller ve ödevler dersin sorumlusu öğretim elemanı tarafından öğrencilere sistem üzerinden çevrimiçi olarak sunulmuştur. Araştırmaya katılan öğrencilerin %61.4'ü erkek, %38,6'sı ise kadın öğrencilerden oluşmaktadır.

### C.Veri Seti

Araştırmada, eğitim ve test için kullanılan veriler 2015-2016 eğitim öğretim yılı güz yarıyılında İnternet Tabanlı Programlama dersini Bulutders üzerinde takip eden öğrencilerin fare tıklaması, girişleri vb. sistemle olan tüm etkileşimlerinden meydana gelmektedir. Öğrencilerin dersteki dönem sonu akademik başarı durumları fakültenin bilgi sisteminden elde edilmiştir.

Veri seti; "Cinsiyet", "Kaç Ödev Gönderdi", "Kaç Gün Sonra Katıldı", "Giriş Sayısı" olmak üzere toplam 6 özelliğten oluşmaktadır. Veri setinde öğrencilerin bu özellikleri ile birlikte sınıflandırma etiketi olarak öğrencilerin dersteki akademik başarı durumları *Başarılı* veya *Başarısız* şeklinde işaretlenmiştir. Kullanılan özellikler ve açıklamalar detaylı olarak Tablo 2.1'de sunulmuştur.

**Tablo 2.1.** Araştırmanın veri setinde kullanılan özellikler ve açıklamaları

Özellik Kodu	Açıklama
CINSIYET	Öğrencinin cinsiyeti
ODEVGONDERMESAYISI	Öğrencinin kaç tane ödev gönderdiği
KACGUNSONRAKATILDI	Öğrencinin Bulutders üzerinde tanımlanan derse kaç gün sonra katıldığı
GIRISSAYISI	Öğrencinin Bulutders'e giriş sayısı
DERSNOTUTIKLAMASAYISI	Öğrencinin DYS üzerinde yer alan ders notlarına kaç kez tıkladığı
BASARIDURUMU	Öğrencinin başarı durumu

Sınıflandırma algoritmalarının eğitimi ve testinde, eğitim için farklı test için farklı veri kümeleri kullanılmaktadır. Diğer bir yol ise bir veri setinin bir kısmının eğitim, bir kısmının ise test için kullanılmasıdır. Bu yöntemlere alternatif olarak kullanılacak bir diğer yöntem ise “*K-Katlamalı Çapraz Doğrulama*” yöntemidir. *K-Katlamalı Çapraz Doğrulama* yönteminde veri seti rastgele  $K$  eşit parçaya bölünerek  $K-1$  parçası eğitim, 1 parçası ise test için kullanılır ve bu işlem  $K$  kez tekrar edilerek algoritmaların ortalama başarısı hesaplanır. Bu yöntem sayesinde özellikle az örnekli veri setleri için aynı veri setinin hem eğitim hem de test işlemi için kullanılabilmesi olanaklı hale gelir. Bu nedenle, bu araştırmada *K-Katlamalı Çapraz Doğrulama* yöntemi  $K=10$  parametresiyle kullanılmıştır.

#### D.Sınıflandırma Algoritmaları

Veri seti üzerinde yapılan analizlerde, Naive Bayes, En Yakın  $k$ -Komşu ve Karar Ağacı sınıflandırma algoritmaları kullanılmıştır. Naive Bayes sınıflandırıcısı, istatistiksel temelli ve oldukça etkin bir sınıflandırma algoritması olup Bayes teoremine dayanır. Buna göre;  $X=\{x_1, x_2, \dots, x_n\}$   $x_i$  özelliklerinden meydana gelen ve sınıfı bilinmeyen veri örneği,  $C_1, C_2, \dots, C_n$  ise sınıf değerleridir. Sınıfı belirlenecek örnek için (1) formülüne göre olasılıklar hesaplanır.

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{P(X)} \quad (1)$$

Formülde  $X$  örneğinin her bir  $x_i$  özelliğinin birbirinden bağımsız olduğu varsayılır ve veri örneğinin her bir özelliğinin olasılık değerleri çarpılarak  $P(X/C_i)$  değeri (2) formülüne göre elde edilir.

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) \quad (2)$$

$P(X)$  değeri tüm sınıflar için sabit olduğundan veri örneğinin hangi sınıfa ait olduğuna karar vermede formülün payında yer alan  $P(X | C_i)P(C_i)$  değeri en yüksek olan sınıf, veri örneğinin sınıfı olarak kabul edilir.

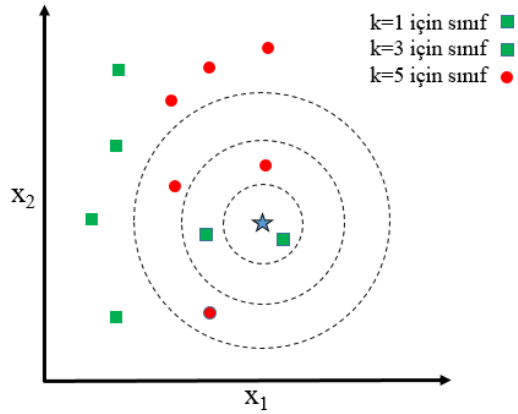
En Yakın  $k$ -Komşu algoritmasında ise, veri seti içerisinde sınıfları belli olan verilerin gözlem değerlerinden yararlanılarak yeni veri örneğinin hangi sınıfa ait olduğu belirlenmeye çalışılır. Sınıfı belirlenmek istenen veri örneğinin, sınıfı belli olan diğer veri örneklerine olan uzaklığı Öklit formülüne göre hesaplanır (3). Uzaklıkları en yakın olan  $k$  adet örnek seçilerek bu örneklerden çoğunluk hangi sınıfa ait ise sınıfı bilinmeyen veri örneğinin sınıfı bu sınıf olarak kabul edilir. Sınıf belirlenirken en yakındaki veri örneklerinin sınıflarının çoğunluğu söz konusu olduğu için  $k$  değeri için genellikle 1, 3, 5 gibi tek sayılar seçilir.

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (3)$$

En Yakın  $k$ -Komşu algoritmasında, seçilen farklı  $k$  değerlerine göre sınıflı belirlenmesi istenen veri örneği farklı sınıflarda yer alabilir. Bu durum Şekil 2.1'de gösterilmiştir.

En Yakın  $k$ -Komşu algoritmasının uygulanma sürecinde veri örneklerinin sınıflarının belirlenmesinde;

- $k$  parametresi belirlenir.
- Sınıflı bilinmeyen veri örneğinin, sınıflı bilinen tüm veri örneklerine olan uzaklıkları Öklit formülüne göre hesaplanır.
- Hesaplanan uzaklıklar sıralanır ve en küçük  $k$  tanesi seçilir.
- Seçilen veri örneklerinde hangisi sınıf daha fazla ise bu sınıf seçilir
- Seçilen bu sınıf bilinmeyen veri örneğinin sınıflı olarak kabul edilir.



**Şekil 2.1** En Yakın  $k$ -Komşu Algoritmasında  $k$  değerlerine göre oluşan farklı sınıflar

Araştırmada kullanılan üçüncü sınıflandırma algoritması olan Karar Ağacı sınıflandırmasında ise sınıf etiketleri bilinen veri örnekleri kullanılarak kural tabanlı bir ağaç yapısı oluşturulmaya çalışılır. Oluşturulan ağaç yapısında, düğümleri veri örneklerinin öz nitelikleri oluşturur. Yapının en üstteki bileşenine *kök*, uçlardaki bileşenlerine *yaprak* ve aradaki diğer bileşenlerine ise *dal* adı verilir. Karar Ağaçları sınıflandırmasında çözülmesi gereken problem, hangi öz niteliğin ağaçta kök olarak yer alacağı ve kökten itibaren düğümlerin dalların ve yaprakların hangi kritere göre oluşturulacağıdır. Bu amaçla, ID3 (J. R. Quinlan, 1986), C4.5 (J Ross Quinlan, 1992), Regresyon Ağaçları (Breiman, Friedman, Olshen, & Stone, 1984) gibi farklı Karar Ağacı algoritmaları oluşturulmuştur. Araştırmada kullanılan C4.5 algoritması entropi tabanlı Karar Ağacı algoritması olarak adlandırılır.

Entropi bir sistemdeki belirsizliğin ölçüsüdür.  $S$  bir kaynak,  $m_1, m_2, \dots, m_n$  bu kaynakta üretilen mesajlar olsun.  $S$  kaynağından  $m_i$  mesajının üretilme olasılığı  $p_i$  ise  $S$  kaynağının entropisi  $H(S)$  (4) formülü ile hesaplanır.

$$H(S) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (4)$$

Entropi temelli Karar Ağacı sınıflandırmasında her bir öznelik için *bilgi kazancı* hesaplanır. Veri örneklerinin özneliklerinden elde edilen bilgi kazanç değerlerine göre ağacın kökü, düğümleri ve yaprakları oluşturulur. Karar Ağaçlarında ağacın büyüklüğü bazı durumlarda çok büyük olabilmektedir. Bu durumda oluşturulan ağacın boyutu, *budama* adı verilen bir yöntemle küçültülür.

#### E.Sınıflandırma Algoritmalarında Sonuçların Değerlendirilmesi

Veri madenciliği araştırmalarında, sınıflandırma algoritmalarının sonuçlarının değerlendirilmesinde karışıklık matrisi üzerinden elde edilen *Doğruluk*, *Hassasiyet*, *Duyarlık*, *Hata Oranı*, *F-Değeri* gibi ölçümler kullanılır. Karışıklık matrisi üzerinde, doğru sınıflandırılmış pozitif sınıf değerleri (DP), yanlış sınıflandırılmış negatif sınıf değerleri (YP), yanlış sınıflandırılmış pozitif sınıf değerleri (YN) ve doğru sınıflandırılmış negatif sınıf değerleri (DN) yer alır. Örnek olarak, gerçekte sınıfı BAŞARILI olan bir öğrenci, algoritma tarafından BAŞARISIZ olarak sınıflandırıldıysa, bu örnek karışıklık matrisinde yanlış pozitif (YP) alanına işlenir. Gerçekte sınıfı BAŞARILI olup algoritma tarafından da BAŞARILI olarak sınıflandırılması durumunda ise doğru pozitif (DP) alanına işlenir. Karışıklık matrisi Tablo 2.2 de detaylı olarak gösterilmiştir.

**Tablo 2.2.** Karışıklık matrisi

Gerçek Sınıf	Öngörülen Sınıf		
		Sınıf 1 (Pozitif)	Sınıf 2 (Negatif)
	Sınıf 1 (Pozitif)	Doğru Pozitif (DP)	Yanlış Negatif (YN)
Sınıf 2 (Negatif)	Yanlış Pozitif (YP)	Doğru Negatif (DN)	

Karışıklık matrisi kullanılarak elde edilen bir ölçüm değeri olan *Doğruluk* ölçümü doğru sınıflandırılmış örnek sayısının toplam örnek sayısına oranıdır ve (5) formülüyle elde edilir. *Hata Oranı* ölçümü ise yanlış sınıflandırılmış örnek sayısının toplam örnek sayısına oranıdır ve formül (6) kullanılarak hesaplanır. *Hata Oranı* ölçümü *Doğruluk* değerinin 1 değerinden çıkarılmasıyla da elde edilebilir.

$$Doğruluk = \frac{DP + DN}{DP + YN + YP + DN} \quad (5)$$

$$Hata Oranı = \frac{YP + YN}{DP + YN + YP + DN} \quad (6)$$

Karışıklık matrisi kullanarak hesaplanan diğer ölçümler ise *Duyarlık* ve *Hassasiyet* ölçümleridir. *Duyarlık* ölçümü doğru sınıflandırılmış pozitif örnek sayısının, doğru sınıflandırılmış pozitif örnek sayısı (TP) yanlış sınıflandırılmış negatif örnek sayısının (YN) toplamına oranıdır ve (7) formülüyle hesaplanır. *Hassasiyet* ölçümü ise doğru

sınıflandırılmış pozitif örneklerin sayısının (DP) toplam pozitif örnek sayısına olan oranıdır ve (8) formülüne göre hesaplanır.

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \quad (7)$$

$$\text{Hassasiyet} = \frac{DP}{DP + YP} \quad (8)$$

*Duyarlılık* ve *Anma* ölçümlerinin harmonik ortalamasının alınması ile elde edilen bir diğer ölçüm ise *F-Değeri* ölçümüdür ve (9) formülüne göre hesaplanır.

$$F - \text{Değeri} = \frac{2DP}{2DP + YP + YN} \quad (9)$$

### F. Verilerin Analizi

Verilerin analizinde, veri madenciliği araştırmalarında kullanılan araçlardan biri olan Weka yazılımı (Frank, Hall, & Witten, 2016) kullanılmıştır. Weka veri setini tanımlamada kendi içerisinde *arrf* dosya uzantılı özel bir format kullanılmaktadır. Bundan dolayı Bulutders veritabanından alınan veriler öncelikle *csv* formatına sonrasında ise *arrf* formatına dönüştürülmüştür. “Ödev Gönderme Sayısı”, “Kaç Gün Sonra Katıldı”, “Giriş Sayısı” ve “Ders Notu Tıklama Sayısı” özellikleri içerisinde yer alan sürekli değerler Weka’nın ayrıklaştırma modülü ile ayrık hale getirilmiştir.

### III. Bulgular ve Yorum

Yapılan analizler sonucunda en yüksek doğruluk değerini %64.3 ile IBk algoritmasının  $k=3$  parametresiyle elde ettiği görülmektedir. IBk algoritmasını J48 Karar Ağacı %62.9 ve yine IBk algoritmasının %57.1 doğruluk değerleriyle  $k=1$ ,  $k=5$  parametreleriyle takip ettiği görülmektedir. Analizler sonucunda algoritmalar tarafından elde edilen doğruluk ve kesinlik değerleri detaylı olarak Tablo 3.1 ile verilmiştir.

**Tablo 3.1.** Sınıflandırma algoritmalarının *doğruluk* ve *kesinlik* değerleri

Algoritma	Doğruluk (%)	F-Değeri
Naive Bayes	55.7	.56
IBk (k=1)	57.1	.57
<b>IBk (k=3)</b>	<b>64.3</b>	<b>.64</b>
IBk (k=5)	57.1	.57
J48	62.9	.63

Naive Bayes algoritmasıyla yapılan analizler sonucunda, karışıklık matrisinde *Doğru Pozitif (DP)*=18, *Yanlış Negatif (YN)*=19, *Yanlış Pozitif (YP)*=12 ve *Doğru Negatif (DN)*=21 olarak hesaplanmıştır. J48 Karar Ağacı algoritmasıyla yapılan analizler sonucunda, karışıklık matrisinde *Doğru Pozitif (DP)*=26, *Yanlış Negatif (YN)*=11, *Yanlış Pozitif (YP)*=15 ve *Doğru Negatif (DN)*=18 olarak gözlemlenmiştir. En yüksek



doğruluk sonucunu veren IBk algoritmasının k=3 parametresiyle yapılan analizler sonucunda, karışıklık matrisinde *Doğru Pozitif (DP)*=25, *Yanlış Negatif (YN)*=12, *Yanlış Pozitif (YP)*=13 ve *Doğru Negatif (DN)*=20 olarak bulunmuştur. Sonuçların detayı Tablo 3.2 ile sunulmuştur.

**Tablo 3.2.** Sınıflandırma algoritmalarına göre karışıklık matrisleri

<i>Algoritma</i>		<i>BAŞARILI</i>	<i>BAŞARISIZ</i>
<i>Naive Bayes</i>	<i>BAŞARILI</i>	18	19
	<i>BAŞARISIZ</i>	12	21
<i>J48</i>	<i>BAŞARILI</i>	26	11
	<i>BAŞARISIZ</i>	15	18
<i>IBk (k=1)</i>	<i>BAŞARILI</i>	19	18
	<i>BAŞARISIZ</i>	12	21
<i>IBk (k=3)</i>	<i>BAŞARILI</i>	25	12
	<i>BAŞARISIZ</i>	13	20
<i>IBk (k=5)</i>	<i>BAŞARILI</i>	23	14
	<i>BAŞARISIZ</i>	16	17

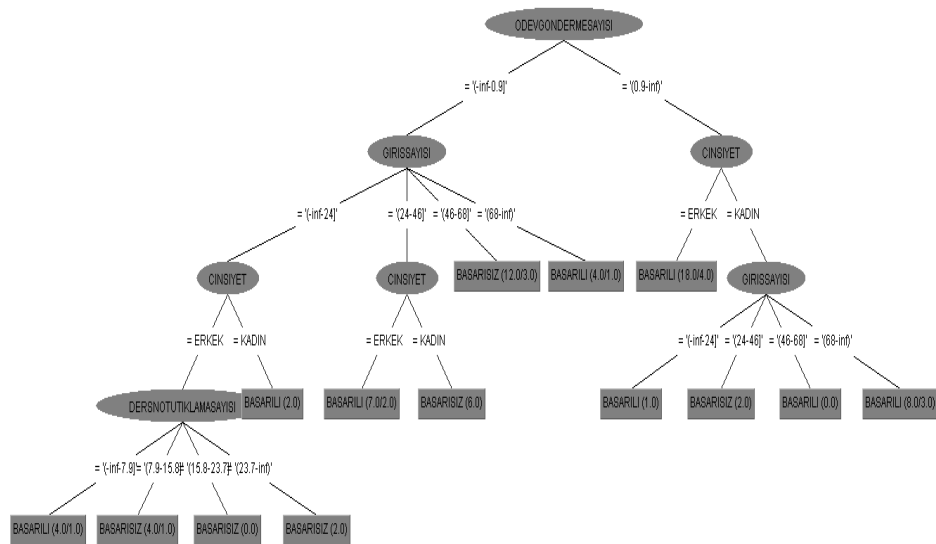
*Naive Bayes* algoritmasının *F-Değeri*, *Hassasiyet* ve *Duyarlılık* sonuçlarına bakıldığında ise *F-Değeri*=0.56, *Hassasiyet*=0.57 ve *Duyarlılık*=0.56 olarak tespit edilmiştir. *J48* Karar Ağacı algoritmasının *F-Değeri*, *Hassasiyet*, *Duyarlılık* sonuçlarına bakıldığında ise *F-Değeri*=0.63, *Hassasiyet*=0.63 ve *Duyarlılık*=0.63 olarak gözlemlenmiştir. *IBk* algoritmasının k=3 parametresiyle *F-Değeri*, *Hassasiyet*, *Duyarlılık* sonuçlarına bakıldığında ise *F-Değeri*=0.64, *Hassasiyet*=0.64 ve *Duyarlılık*=0.64 olarak bulunmuştur. Detaylı sonuçlar Tablo 3.3 ile verilmiştir.

**Tablo 3.3.** Sınıflandırma algoritmaları ile elde edilen *F-Değeri*, *hassasiyet* ve *duyarlılık* değerleri

<i>Algoritma</i>	<i>Sınıf</i>	<i>F-Değeri</i>	<i>Hassasiyet</i>	<i>Duyarlılık</i>
<i>Naive Bayes</i>	<i>BAŞARILI</i>	.54	.60	.49
	<i>BAŞARISIZ</i>	.58	.53	.64
	<i>Ağırlıklı Ortalama</i>	.56	.57	.56
<i>J48</i>	<i>BAŞARILI</i>	.68	.63	.70
	<i>BAŞARISIZ</i>	.58	.62	.55
	<i>Ağırlıklı Ortalama</i>	.63	.63	.63
<i>IBk (k=1)</i>	<i>BAŞARILI</i>	.56	.61	.56
	<i>BAŞARISIZ</i>	.58	.54	.58

	Ağırlıklı Ortalama	.57	.58	.57
IBk (k=3)	BAŞARILI	.67	.66	.68
	BAŞARISIZ	.62	.63	.61
	Ağırlıklı Ortalama	.64	.64	.64
IBk (k=5)	BAŞARILI	.61	.59	.62
	BAŞARISIZ	.53	.55	.52
	Ağırlıklı Ortalama	.57	.57	.57

Karar Ağacı algoritması tarafından oluşturulan modelin ağaç yapısında kök düğümünü “Ödev Gönderme Sayısı” özelliğinin, bir alt düğümleri ise “Cinsiyet” ve “Giriş Sayısı” özelliklerinin oluşturduğu görülmektedir. Oluşan modelin ağaç yapısının tümü Şekil 3.1’de gösterilmiştir.



Şekil 3.1. J48 Karar Ağacı algoritması tarafından oluşturulan modelin ağaç yapısı

Bu çalışmada Algoritmalar tarafından öğrencilerin akademik başarılarının tahmin edilmesinde, DYS yazılımı ile olan hangi etkileşimlerin ne kadar bilgi kazancı sağladığıyla ilgili olarak yapılan analizlerde; veri seti içerisinde yer alan her bir özelliğin bilgi kazancı değerleri hesaplanmıştır. Tablo 3.4’de görüldüğü gibi, en yüksek bilgi kazancını “Ödev Gönderme Sayısı” özelliğinin sunduğu bunu sırasıyla “Giriş Sayısı” ve “Kaç Gün Sonra Katıldı” özellikleri takip etmektedir. “Cinsiyet” özelliğinin ise en az bilgi kazancı sağladığı ortaya çıkmıştır.

**Tablo 3.4.** Veri setindeki özelliklerin bilgi kazanç değerleri

<i>Özellik</i>	<i>Bilgi Kazancı</i>
<i>ODEVGONDERMESAYISI</i>	<i>.054</i>
<i>GIRISSAYISI</i>	<i>.045</i>
<i>KACGUNSONRAKATILDI</i>	<i>.043</i>
<i>DERSNOTUTIKLAMASAYISI</i>	<i>.028</i>
<i>CINSIYET</i>	<i>.013</i>

#### **IV. Tartışma ve Sonuç**

Araştırma sonucunda elde edilen bulgulara bakıldığında, sınıflandırma algoritmalarının doğruluk değerlerinin %55.7 ile %64.3 arasında değiştiği görülmektedir. En yüksek doğruluk değerini %64.3 ile En Yakın k-Komşuluk algoritması k=3 parametresiyle elde etmiştir. Osmanbegović ve Suljić (2012) gerçekleştirdikleri çalışmada öğrencilerin başarı durumlarının tahmininde, algoritmaların doğruluk değerlerini %71.2 ile %76.6 değerleri arasında bulmuşlar ve çalışmalarında en yüksek doğruluk değeri Naive Bayes algoritması ile elde edilmiştir. Akçapınar vd. (2015) tarafından yapılan diğer bir çalışmada, öğrencilerin çevrimiçi öğrenme ortamındaki etiket kullanma sayısı, ödev sayısı, gezinti sayısı gibi özelliklerden oluşan veri setinde Naive Bayes algoritması ile %75.4 doğruluk değeri elde edilmiştir.

Alanyazındaki araştırmaların sonuçlarıyla karşılaştırıldığında, bu çalışmada elde edilen doğruluk değerlerinin, söz konusu araştırmalarda elde edilen değerlerin biraz altında olduğu görülmektedir. Osmanbegović ve Suljić (2012) çalışmasını 257 örnek ve 11 özellik bulunan bir veri seti kullanarak gerçekleştirmiştir, Akçapınar vd. (2015) ise çalışmalarında 76 örnek ve 10 özellik bulunan bir veri seti kullanmışlardır. Bu çalışmada ise 70 örnek ve 6 özelliğinden oluşan bir veri seti kullanılmıştır. Daha az özelliğe sahip olan veri kümelerinde, veri seti üzerindeki işlemler için daha az bilgi işlem kaynağı tüketilmekte ve algoritmalar daha hızlı sonuç vermektedir. Bu çalışmadaki elde edilen değerler ve diğer araştırmalardaki değerlerle karşılaştırıldığında, sınıflandırma algoritmaları tarafından öğrencilerin akademik başarılarının tahmin edilmesinde örnek ve özellik sayısının artırılması ile sınıflandırma algoritmalarının daha yüksek doğruluk değerine ulaşacağı şeklinde düşünülebilir.

Öğrencilerin akademik başarılarının tahmin edilmesinde, sınıflandırma algoritmalarına DYS yazılımı ile olan hangi etkileşimlerin ne kadar bilgi kazancı sağladığı ile ilgili olarak elde edilen bulgulara göre; veri setinde yer alan “*Ödev Gönderme Sayısı*” özelliğinin .054 bilgi kazancı değeri ile ilk sırada yer aldığı görülmektedir. Bu özellik, veri seti içerisinde DYS sistemini kullanan öğrencilerin, sistem üzerinden verilen ödevlerle olan etkileşimini sayısal değer olarak yansıtmaktadır. Akçapınar vd. (2015) yaptıkları çalışmada veri setinin özelliklerini sıraladıklarında, öğrencinin sisteme giriş sayısını en üst sırada yer alan özellik olarak gözlemlemiştir. Osmanbegović ve Suljić (2012)’in araştırmalarında ise özellik sıralamasında, sosyodemografik niteliğe sahip olanların üst sıralarda olduğu görülmektedir. Bhardwaj (2011)

tarafından yapılan araştırmada ise öğrencilerin orta öğretimdeki başarı notunun özellik sıralamasında en üstte yer aldığı görülmektedir. Bu araştırma ile eğitim öğretim faaliyetleri içerisinde sıklıkla yer alan ve öğrencileri değerlendirmede önemli bir role sahip olan ödevlerin, sınıflandırma algoritmalarının öğrencilerin akademik başarılarının tahmininde önemli bir role sahip olabileceği ve öğrencilerin DYS sistemi üzerinden verilen ödevlerle olan etkileşimlerinin de sınıflandırma algoritmalarında en yüksek sırada yer alabileceği ortaya konmuştur.

Bu araştırmanın alanyazına sağladığı önemli bir katkı, öğrencilerin kişisel, sosyo-demografik özelliklerinin yanısıra eğitim öğretim faaliyetleri içerisinde kullanılan çevrimiçi yazılımlarla olan etkileşimlerinin de, akademik başarılarını tahmin etmede alternatif bir yol olarak kullanılabilmesinin göstermiş olmasıdır. Sonuç olarak, sözkonusu yazılımların ders notu paylaşımı, ödev yönetimi, çevrimiçi öğrenme ortamı sağlaması gibi öğretimsel faydalarının yanısıra, içerisine entegre edilebilecek modüllerle eğitim öğretim planlamacılarına, öğreticilere ve eğitim yöneticilerine öğrencilerin akademik performanslarını tahmin etmede de önemli bir katkı sunabilir.

#### Kaynaklar

- Ahmad, F., Ismail, N. H. ve Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9(129). <https://doi.org/10.12988/ams.2015.53289>
- Akçapınar, G., Altun, A. ve Aşkar, P. (2015). Modeling Students' Academic Performance Based on Their Interactions with the Online Learning Environment. *İlköğretim Online*, 14(2), 815–824. <https://doi.org/10.17051/io.2015.03160>
- Akçapınar, G., Coşgun, E. ve Altun, A. (2013). Mining Wiki Usage Data for Predicting Final Grades of Students. *Prague 2013 (IAC-ETeL 2013)*, 1–6.
- Baldi, P., ve Brunak, S. (2001). *Bioinformatics - The machine learning approach. Machine Learning*.
- Baradwaj, B. ve Pal, S. (2012). Mining educational data to analyze student's performance. *Internation Journal Od Advamced Computer Science and Applications*, 2(6), 63–69. <https://doi.org/vol.2,No.6>
- Bhardwaj, B. K. (2011). Data Mining : A prediction for performance improvement using classification. *(IJCSIS) International Journal of Computer Science and Information Security*, 9(4).
- Breiman, L., Friedman, J. H., Olshen, R. A., ve Stone, C. J. (1984). *Classification and Regression Trees. Wadsworth International Group*. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Classification+and+regression+trees#0>
- Buniamin, N., Mat, U. Bin ve Arshad, P. M. (2016). Educational data mining for prediction and classification of engineering students achievement. In *IEEE 7th International Conference on Engineering Education, ICEED 2015*. <https://doi.org/10.1109/ICEED.2015.7451491>
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. ve Zanasi, A. (1998). *Discovering data mining : from concept to implementation*. Prentice Hall.

- Christian, T. M. ve Ayub, M. (2014). Exploration of classification using NBTree for predicting students' performance. In *Proceedings of 2014 International Conference on Data and Software Engineering, ICODSE 2014*. <https://doi.org/10.1109/ICODSE.2014.7062654>
- Del, A. (2016). Predicting academic performance in traditional environments at higher-education institutions using data mining : A review, (February 2017).
- Fayyad, U. M., Weir, N. ve Djorgovski, S. G. (1993). Automated Cataloging And Analysis Of Sky Survey Image Databases: The Skicat System. *Cikm*, 527–536. <https://doi.org/10.1145/170088.170414>
- Frank, E., Hall, M. A. ve Witten, I. H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"; *Morgan Kaufmann, Fourth Edition*.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. ve Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–4. <https://doi.org/10.1038/nature07634>
- Güldal, H. (2015). Servis Olarak Bir Ders Yönetim Sistemi Yazılımı: Bulutders.com. In *5th International Conference on Research in Education – ICRE*. Edirne.
- Márquez-Vera, C., Cano, A., Romero, C. ve Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied Intelligence*, 38(3), 315–330. <https://doi.org/10.1007/s10489-012-0374-8>
- Martins, L. L. ve Kellermanns, F. W. (2004). A Model of Business School Students' Acceptance of a Web-Based Course Management System. *Academy of Management Learning & Education*, 3(1), 7–26. <https://doi.org/10.5465/AMLE.2004.12436815>
- Mishra, T., Kumar, D. ve Gupta, S. (2014). Mining students' data for prediction performance. *International Conference on Advanced Computing and Communication Technologies, ACCT*, 255–262. <https://doi.org/10.1109/ACCT.2014.105>
- Osmanbegović, E. ve Suljić, M. (2012). Data mining approach for predicting student performance. *Journal of Economics and Business*, X(1), 3–12.
- Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1(1), 81–106. <https://doi.org/10.1023/A:1022643204877>
- Quinlan, J. R. (1992). *C4.5: Programs for Machine Learning*. Morgan Kaufmann San Mateo California (Vol. 1). [https://doi.org/10.1016/S0019-9958\(62\)90649-6](https://doi.org/10.1016/S0019-9958(62)90649-6)
- Shahiri, A. M., Husain, W. ve Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, 72, 414–422. <https://doi.org/10.1016/j.procs.2015.12.157>
- Tüzüntürk, S. (2010). Veri Madenciliği ve İstatistik, (2001), 65–90.